

Plausibility and causal claims in theory-based studies and evaluation

Spring Conference 2019 of the Working Group “Methods in Evaluation”

DeGEval – Evaluation Society

Causality and Plausibility – Evaluation between Science and Practice

Vienna, Austria, June 20-21, 2019

Derek Beach, PhD

Professor

Department of Political Science

University of Aarhus, Denmark

Email: derek@ps.au.dk



Plausibility and causal inferences in evaluation

- Bayesian logic = updating of confidence based on new evidence
- degrees of confidence (degree of plausibility based on evidence)

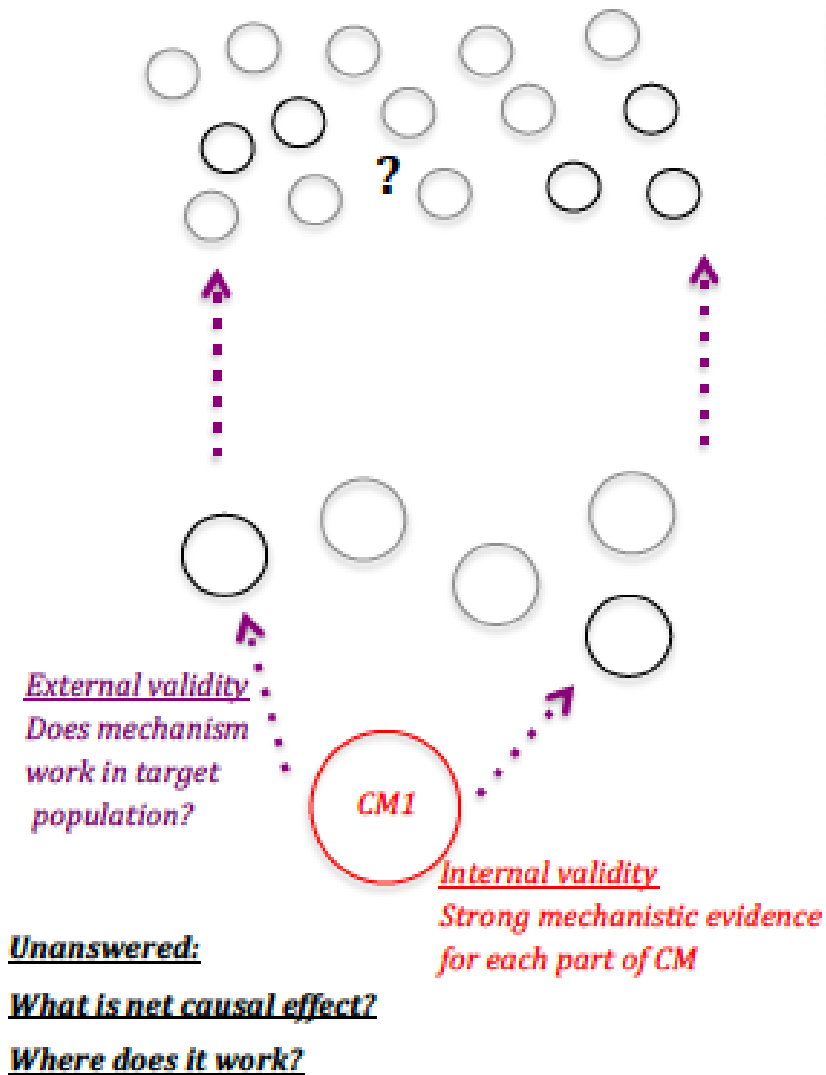
Linguistic form	Numerical equivalent
<i>Certainty, no question about</i>	<i>1.0 (100%)</i>
<i>Almost certainly, beyond reasonable doubt</i>	<i>0.9 (90%)</i>
<i>Very probably</i>	<i>0.8 (80%)</i>
<i>Probably</i>	<i>0.7 (70%)</i>
<i>On balance, somewhat more likely than not</i>	<i>0.6 (60%)</i>
<i>Like as not, even money</i>	<i>0.5 (50%)</i>
<i>Somewhat less than even chance</i>	<i>0.4 (40%)</i>
<i>Probably not</i>	<i>0.3 (30%)</i>
<i>Very probably not</i>	<i>0.2 (20%)</i>
<i>Almost certainly not</i>	<i>0.1 (10%)</i>
<i>Certainly not, impossible</i>	<i>0.0 (0%)</i>

The argument

- case-based, mechanism-focused evaluation makes fundamentally different types of causal claims, evidenced using very different empirical material
- two parallel evidence hierarchies for causal inferences
- RCT = *counterfactual, cross-case* claims about *net causal effects*
- case-based = *causal mechanisms* as they *operate within cases*
- both have challenges regarding external validity (extrapolation)

Case-based (bottom-up) approach

'how it actually works'



Variance-based (top-down) approach

'it works somewhere'

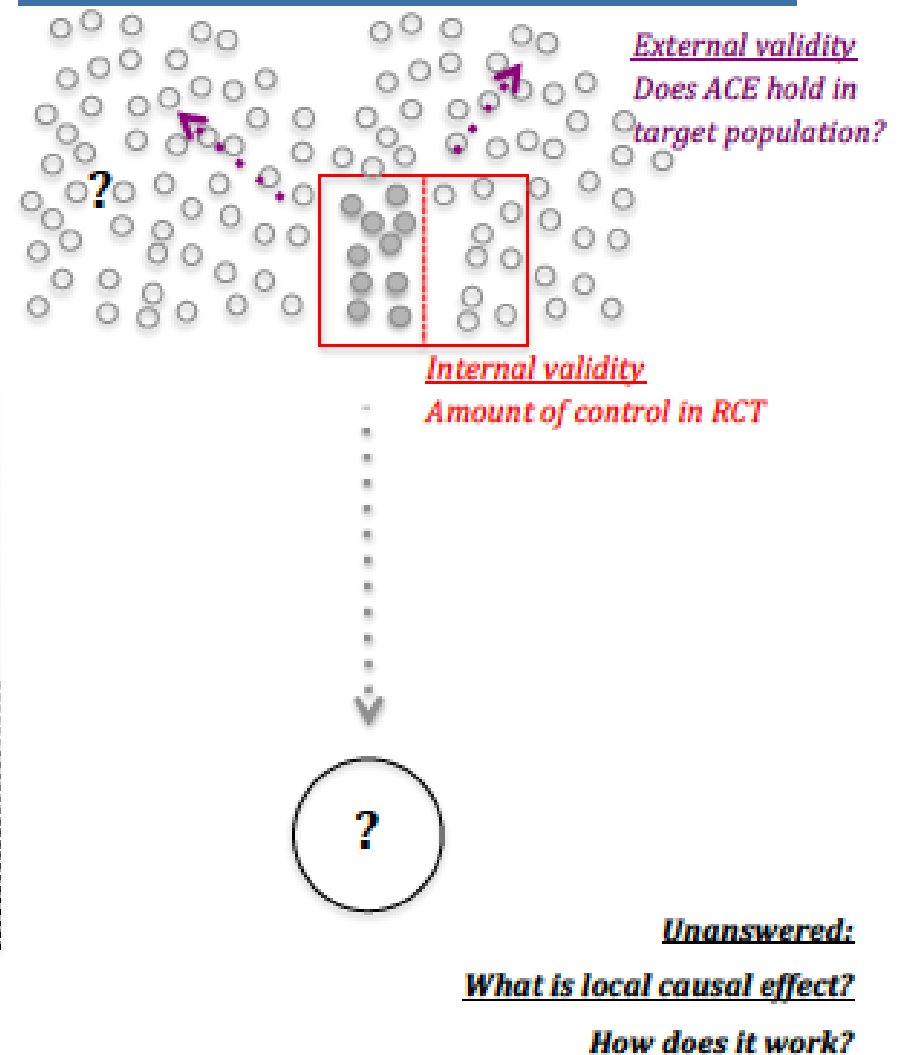


Figure - Case-based versus variance-based evaluation approaches

<p><u>Variance-based</u> Cross-case, counterfactual claims about causal effects of X on Y</p>	<p><u>Case-based</u> Within-case, mechanistic claims about process linking cause and outcome</p>	
Meta-study (lab experiments)	<i>external validity</i>	Multiple PT case studies in population
Lab experiment	<i>internal validity</i>	PT case study
Field experiment		Non-mechanistic case study
Natural experiment		Small-n comparison
Large-n observational		Medium-n comparison
Small-n observational		

Two parallel evidence hierarchies

Case-based evaluation methods

- 'case' = an instance of an intervention (cause) → outcome
- theory defines level of analysis, time period, etc...
- causes always work WITHIN cases (also in RCT)

Case-based evaluation methods

- evaluation case study methods : contribution analysis, theory of change
- problems:
 - links between parts of process treated as 'assumptions'
 - unclear about what evidence actually is
- Process-tracing (PT) more explicit on both dimensions
- see Wauters and Beach (2018), Schmitt and Beach (2015)...

Plausibility and Process-tracing

- internal plausibility = strength of inference about causal mechanism in case
 1. theoretical (productive continuity with activities linking parts made clear)
 2. empirical (strong mechanistic evidence for all activities)
- external plausibility = does it work in other cases?
 1. strategic probing of target population to see whether it works in same fashion

Internal plausibility I – theorized mechanism

- causal mechanism is NOT a counterfactual
- counterfactual (RCT)

...if the first object had not been, the second never had existed.' (Hume, 1927: 157)

- claim that X was the cause of Y based upon studying whether the *absence* of X results in the *absence* of Y, all other things being held equal
- counterfactual does not tell us HOW something works...

Internal plausibility I – theorized mechanism

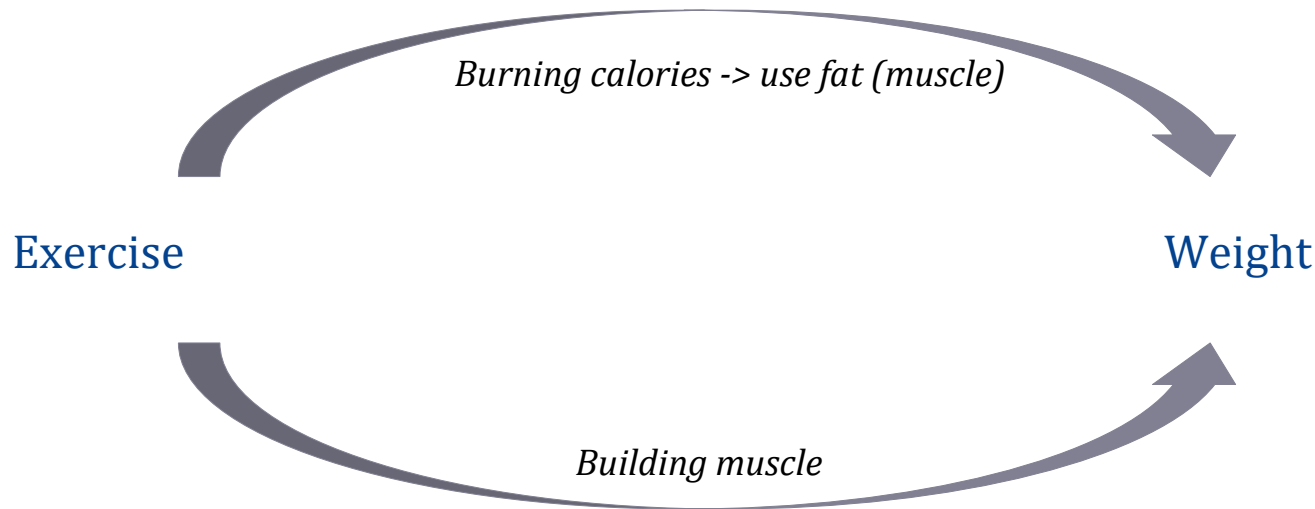
Causality = causal mechanisms

- Open up ‘black box’ between cause (C) and outcome (O)
- how process actually works
- Study the dynamic, interactive influence of causes upon outcomes, and in particular how *causal forces* are *transmitted* through a series of interlocking parts of a causal mechanism to produce an outcome.

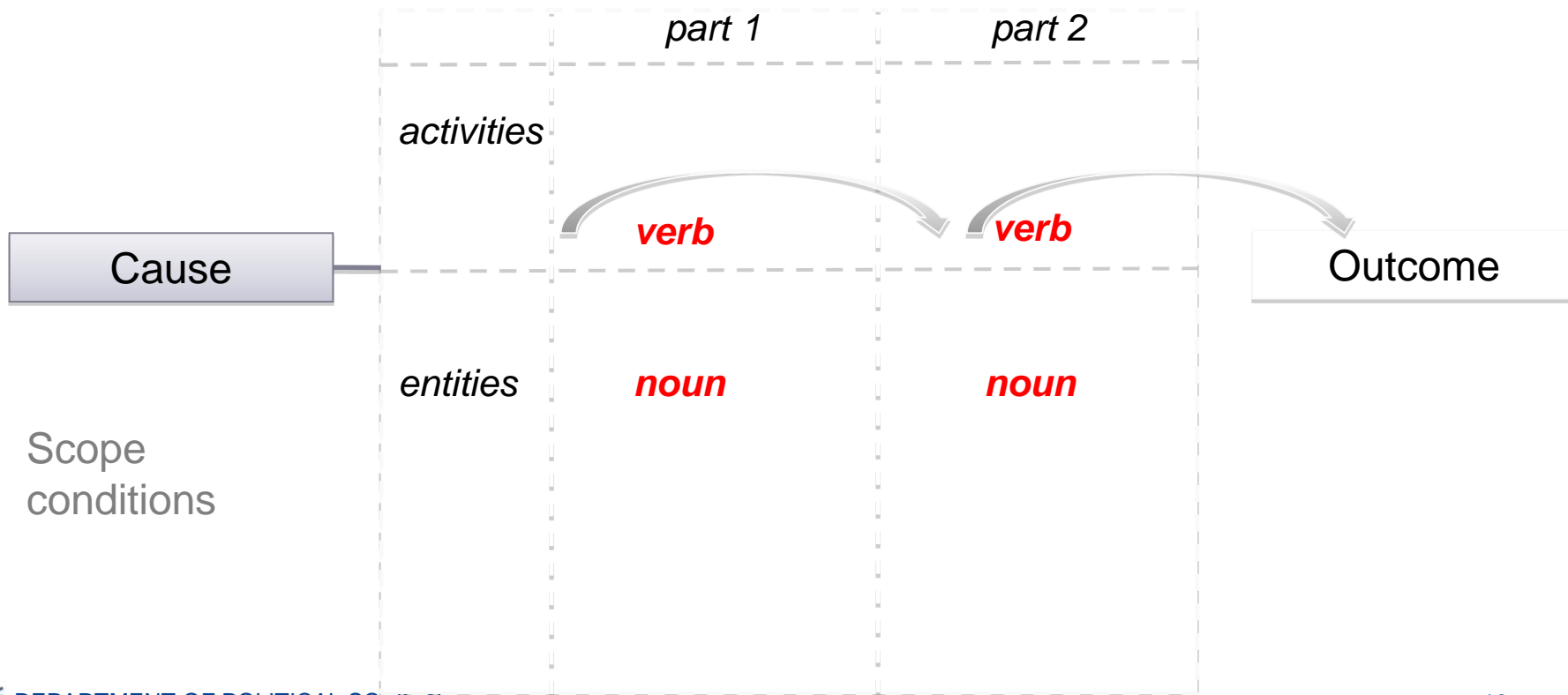
Internal plausibility I – theorized mechanism

The problem of 'masking' – cannot assess net causal effect with PT

- complex interventions trigger multiple mechanisms that can link to SAME outcome



Internal plausibility I – theorized mechanism



Internal plausibility I – theorized mechanism

Cause	Part 1	Part 2	Outcome
Educational program in nutrition for mothers	Mother attends courses and learn about nutrition	Mother uses acquired knowledge to: a) purchase better food b) prepare better food	Better child nutrition

Source : White 2009

Internal plausibility I – theorized mechanism

To qualify as a mechanistic explanation = explains 'how it works'

- productive continuity = causal logic binding parts together made explicit
- activities – what entities are doing

Superficial mechanistic explanation	Cause	→ 'one-liner' description of process →					Outcome
Incomplete mechanistic explanation	Cause	→	entity	→	entity	→	Outcome

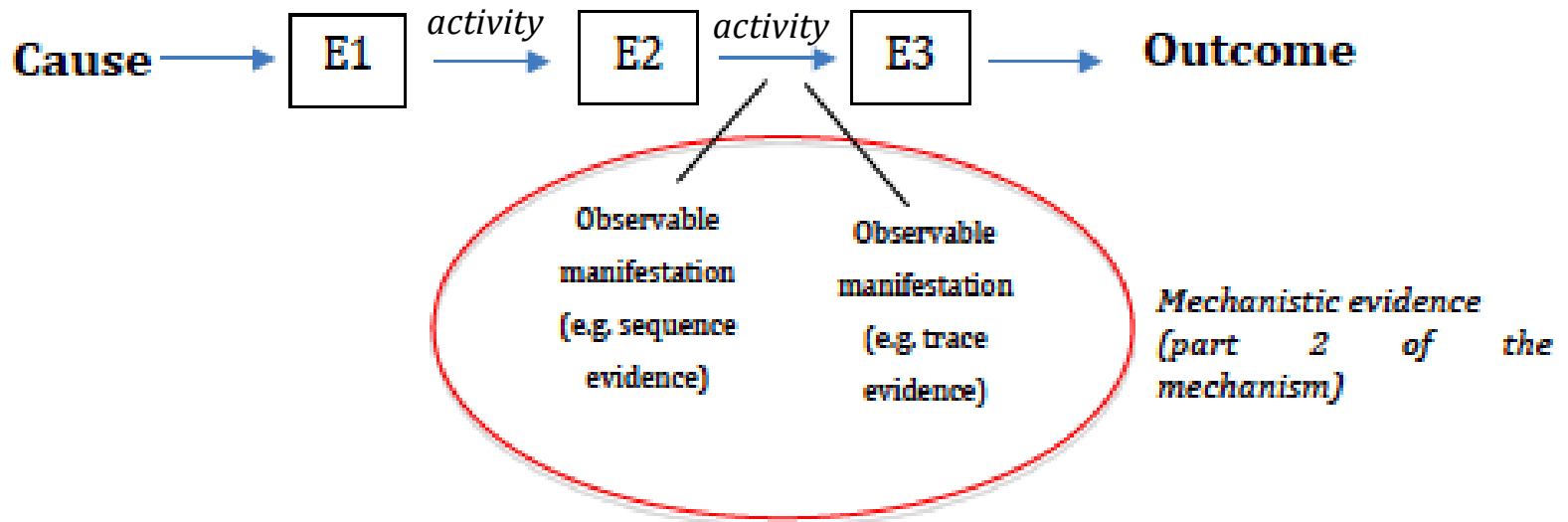
Table 2.2 – Incomplete mechanistic explanations - superficial and incomplete

Source: adapted from Craver and Darden, 2013: 83-95.

Internal plausibility II – mechanistic evidence

Mechanistic evidence

- activities should leave observational traces in actual cases, not just compelling story...



Internal plausibility II – mechanistic evidence

- 4 types of mechanistic, within-case evidence:
 1. Pattern
 2. Sequence
 3. Trace
 4. Account

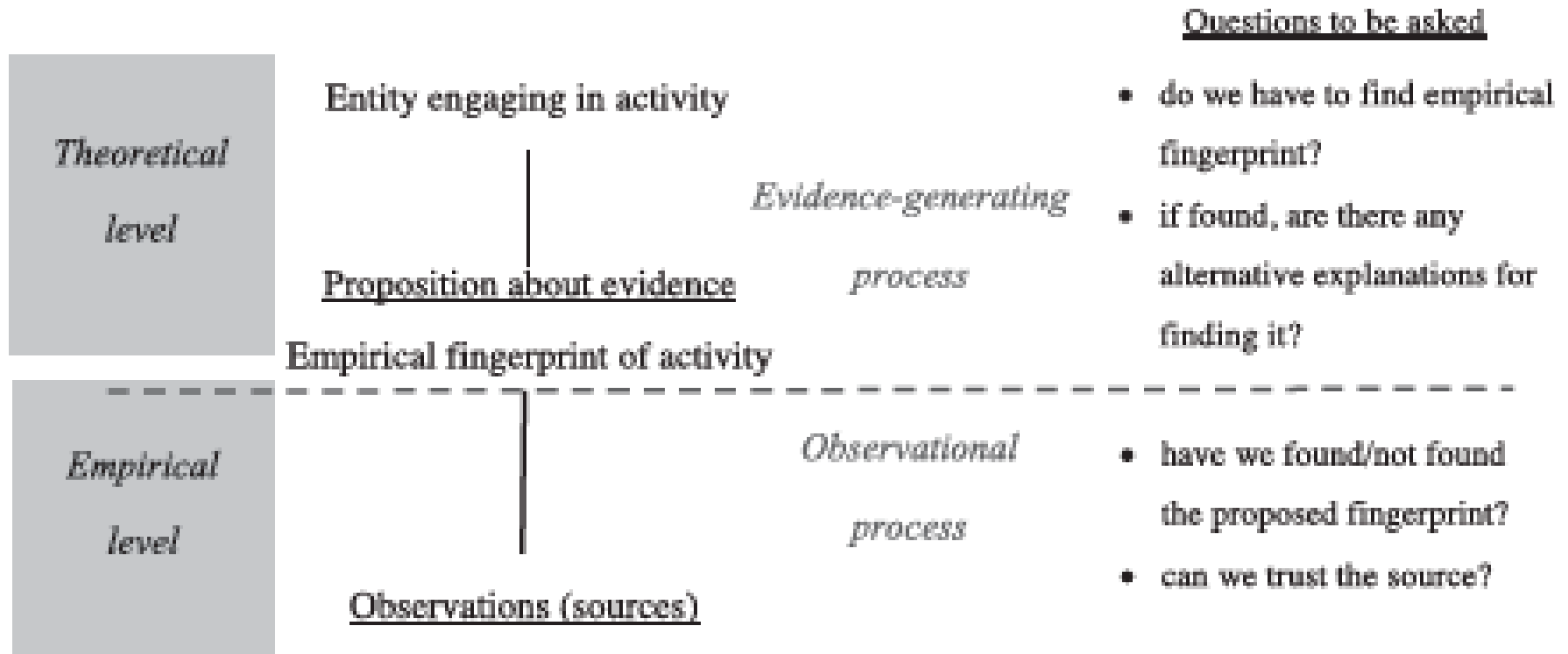


Fig. 5.1. A two-stage evidence-evaluation framework for turning empirical material into evidence of mechanisms

Internal plausibility II – mechanistic evidence

Mechanistic evidence

- theoretical certainty -> what we have to find
- disconfirming power of evidence
- theoretical uniqueness -> if found, are there alternative explanations
- confirming power of mechanistic evidence
- ideally direct evidence of activity that cannot be accounted for with alternative explanations of evidence



1. Rational employee mechanism

Cause		Mechanism				Outcome
		<i>Part 1</i>	<i>Part 2</i>	<i>Part 3</i>	<i>Part 4</i>	
<p>HR manager and training manager of a company prepare a training project. After a job task analysis and organization analysis they enquire after the needs of their employees. A person analysis is being performed.</p>	<p><i>Entities and activities</i></p>	<p>The employees receive the questions of their managers. They reflect on how they are doing their job, what needs they perceive and submit their answers/indicate their needs.</p>	<p>The HR service examines the received need indications and propose the training that seems to be a good option for the employees and the company in general (probably a compromise).</p>	<p>The employees receive the proposal from their managers. They weigh of the pros (new knowledge, better performance, good image, ...) and cons (time management, effort,) of following the training.</p>	<p>If the outcome of the rational process of overthinking the training is positive, the employees follow the training. They are motivated, as they understand their need.</p>	<p>Effective training transfer. Back at work the employees apply the new skills, technique on the job.</p>
			<p>At this point, there might be a feedback loop between part 2 and part 3. The employees might not be completely convinced. The management might f.e. need to explain the need for training better, the goal,</p>			

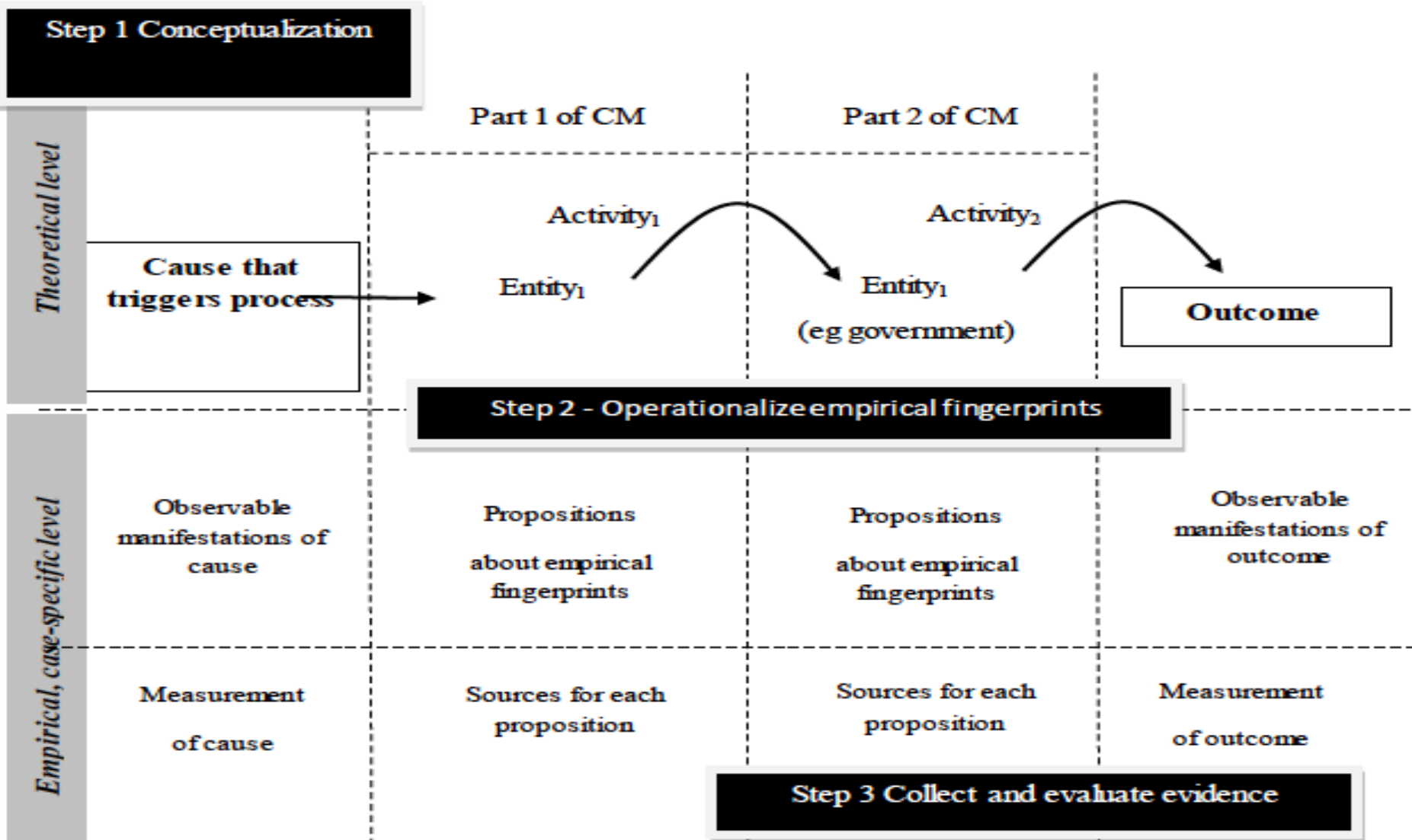
Rational employee's reflections about job

- predicted mechanistic evidence = written analysis by employee (e.g. notes)
 - low certainty (not have to find)
 - but if find, relatively unique (other explanations for finding analysis not plausible)

Therefore : if find e (and can trust it), strong updating in confidence that P1 took place

Mechanism
<i>Part 1</i>
The employees receive the questions of their managers. They reflect on how they are doing their job, what needs they perceive and submit their answers/indicate their needs.

Internal plausibility in Process-tracing



Standards for internal plausibility

	High plausibility	Low plausibility
Theoretical mechanism	Productive continuity with unbroken chain of activities	Black-boxed mechanism (minimalist)
What evidence in theory tells us	Direct and unique mechanistic evidence (smoking guns)	Indirect, low uniqueness evidence (straw-in-the-wind)
Actual evidence	Strong sources and full access to empirical record	Weak sources and/or lack of access to empirical record

External plausibility – extrapolation

Mechanistic heterogeneity

- PT sheds some light on 'under what conditions', but...
- same intervention can work differently in different context
- comparison at level of causes/outcome tells us little about mech. heterogeneity

Case	O	C1	C2	C3	C4	UC	Mechanism Operative	Comment
4	1	1	1	0	0	0	C1*C2 \mapsto part 1a \mapsto part 2a \mapsto part 3a \mapsto O	studied with initial process-tracing study
5	1	1	1	0	0	1	C1*C2*UC \mapsto part 1 \mapsto part 2 \mapsto part 3a \mapsto part 3b \mapsto O	mechanism differs at the final stage
3	1	1	1	0	1	0	C1*C2*C4 \mapsto part 1c \mapsto part 2a*2c \mapsto part 3a*3c \mapsto O	mechanism differs at every stage

Table 4.4. – Mechanistic heterogeneity hidden behind known and unknown omitted conditions.

External plausibility – extrapolation

Mechanistic heterogeneity

- have to empirically validate that same mechanism throughout target population
- snowballing-outwards strategy – most similar to most different cases
- can be made easier in other cases by:
 1. focusing on key part of CM
 2. focusing on unique signature of part of CM
- if differences – split population into sub-populations

Standards for external plausibility

- high external plausibility = strategically explored impact of context and determined that no differences within target population
- lower external plausibility = only assessed 1-2 other cases

<p><u>Variance-based</u> Cross-case, counterfactual claims about causal effects of X on Y</p>	<p><u>Case-based</u> Within-case, mechanistic claims about process linking cause and outcome</p>	
Meta-study (lab experiments)	<i>external validity</i>	Multiple PT case studies in population
Lab experiment	<i>internal validity</i>	PT case study
Field experiment		Non-mechanistic case study
Natural experiment		Small-n comparison
Large-n observational		Medium-n comparison
Small-n observational		

Two parallel evidence hierarchies

Take homes

- Process-tracing evaluation methods asks very different questions that are evidenced using very different empirical material
- Level of plausibility (internal/external) required depends on evaluation situation
- Maximalist : high internal and high external
- Explorative (plausibility probes) : low internal and no real external
- Typical : high internal and moderate external

Take homes

- RCT and PT case studies should be used together in larger-scale evaluation projects
 1. RCT = net causal effects
 2. PT = how it works
 3. PT = better design of intervention (what is a cause)
 4. PT = under what conditions

Further reading

- Wauters and Beach (2018) Process tracing and congruence analysis to support theory based impact evaluation. *Evaluation* 24(3): 284–305.
- Beach and Pedersen (2019) *Process-tracing methods*. Ann Arbor: University of Michigan Press.
- Clarke, Gillies, Illari, Russo, and Williamson (2014). Mechanisms and the Evidence Hierarchy. *Topoi* 33 (2): 339–60.